

AD 709522

Submitted to Psychological Monographs 153-259  
Rec'd 1 Apr 9 3 65

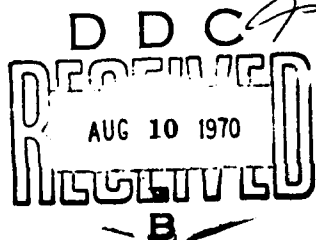
An Analysis of Performance Goals and their Relationship  
to Level of Performance

Edwin A. Locke and Judith F. Bryan

American Institutes for Research  
Washington Office

Abstract

Two experiments were carried out in which goal instructions were used in an attempt to produce variation in performance goals and thus performance level. Factorially independent performance tasks (where relatively little learning was possible) were used in the two experiments. No significant a priori experimental effects were found in either case, apparently due to the fact that most subjects were not pursuing their assigned goals. When subjects were ~~re~~-grouped according to their stated a posteriori performance goals, highly significant goal effects were found. In general, a high or hard goal led to the highest performance level, a low or easy goal to the lowest level, and the goal of "do best" to an intermediate level, though closer to the high than to the low level. "Do best" goals also led to the highest within-group variation in performance. The effect of the goal to "improve" depended on the rate of improvement tried for. The implications of the results for motivation research are discussed.



This document has been approved  
for public release; no sale; its  
distribution is unlimited.

An Analysis of Performance Goals and their Relationship  
to Level of Performance<sup>1</sup>

Edwin A. Locke and Judith F. Bryan

American Institutes for Research  
Washington Office

Since Mace's studies of performance goals in the 1930's, the use of performance goals as independent variables in psychological research has been almost non-existent. Reviews of the extensive literature on level of aspiration indicate that it has been used almost exclusively as a dependent variable (e.g., see Festinger, 1942; Frank, 1941; Fryer, 1964; Lewin, 1958; Rotter, 1942). In the few studies where the actual level of the aspired goal was used as an independent variable (e.g., Locke, 1965; Seigal and Fouraker, 1960) highly significant (positive, linear) relationships were found between the level of the goal and actual level of performance.

There have also been a few studies on the effects of qualitatively different goals on performance. The studies by Mace (1935) referred to above found that the highest level of performance on a mathematical learning task involving computation was achieved by subjects who were given specific and reasonably hard performance goals each day based on their initial skill. The goal of "doing as well as possible" (or "do best" for short) was generally found to result in a lower level of performance than a hard standard but in a higher level of performance than a goal of "improving over one's previous performance." Locke and Bryan (1965b) using the same task as Mace, also found that hard performance standards resulted in the highest level of performance, with the "do best" goal second best. The goal to "improve" was found to result in generally lower performance as in the Mace studies.

Locke and Bryan (1965a) also found hard performance goals to result in a significantly higher level of performance than the goal of "do best" on a psychomotor learning task.

Even less attention has been paid the effects of various types of "low" performance goals. Mace (1935) found that the lowest rate of learning was achieved by subjects given a constant performance goal (one which most subjects reached about halfway through the experiment). Locke and Bryan (1965b) found that miscellaneous goals such as "tried for accuracy," "tried to keep going," or "had no particular goal except to get through" resulted in the lowest level of performance, as compared to other goals.

The present studies were designed to build on this previous work. The major abstract purpose was to illustrate the practical utility of using performance goals as independent variables. At the specific level it was decided to utilize performance tasks (i.e., where little improvement is to be expected with practice) in the present experiment rather than learning tasks. Only one previous study in this area had used performance tasks (Locke's 1965 study, where the tasks involved things like giving uses for objects), and these were tasks where some learning actually was found to occur (also only quantitatively different goals were used in that study). One difference that might be expected between the effects of the goals to "improve" and "do best" on a performance as opposed to a learning task might be that there should not be as much, if any, superiority of the "do best" goal to the "improvement" goal on a performance task. In a learning task some "improvement" is taken for granted and relatively easy to achieve, and a goal simply to "improve" leaves open to question of "how much?" improvement the person will try for and "at what rate?" Thus a goal of improvement could actually retard performance over the goal of "do best"

on a learning task, as seems to have been the case in the studies of Mace (1935) and Locke and Bryan (1965b) cited above. However, on a performance task where improvement is difficult if not impossible, the goal of "improvement" might be expected to compare favorably with the goal of "do best." The present studies were designed to test this possibility.

A third purpose of the present studies was to delineate and classify the various types of goals more systematically than had been done previously (e.g., Locke and Bryan, 1965b). Previous work by the present investigators had made it possible to develop goal questionnaires on which subjects could check the goal which they were pursuing rather than having to explain their goals in their own words (which can lead to problems of misunderstanding of the concept of "goal" and occasional problems of interpretation).

Two final purposes of the present studies were to determine how likely subjects are to accept different types of goal instructions and to examine the effects of extrinsic incentives on goal-setting activity.

In both the studies to be reported here we chose to use actual ability tests as performance tasks. One reason for this, in addition to the desire to achieve generality over tasks, was that such tasks provide the maximum amount of challenge to the investigator who wishes to relate performance goals to performance level. Since such tests are usually thought to measure "maximal" performance, little variation from trial to trial (except for "random" error or subject unreliability) is to be expected. Positive results in this case should be least "obvious." For this same reason, and also to enable us to look at performance over time, short (two-minute) trial periods were used in both experiments; thus fatigue and individual differences in persistence, which would be likely to be quite sensitive to different goals over the long term would not be likely to have much effect.

The goal instructions used for the two experiments were designed to produce the maximum possible variety of goals, both quantitatively and qualitatively. Special attention was devoted to producing low motivation goals which had received relatively little attention previously. It was not expected that all subjects would necessarily follow their assigned goals, but it was hoped that at least wide individual differences in goals would result. The main difference in goal instructions between the two studies was that in the second study two of the three groups of subjects who were not assigned specific quantitative goals were required to set specific levels of aspiration for each trial so that the exact level at which they set their goals could be determined.

#### Experiment I

##### Method

##### Subjects

The subjects were 72 paid college student volunteers (47 males and 25 females) recruited from five Washington, D. C. area colleges during the summer session.

##### Task

The task was a number crossing test developed by Thurstone and described by Moran and McFerd (1959), as a measure of perceptual speed and was similar in nature to the tests used to measure Factor P in French, Ekstrom, and Price's (1963) reference kit for cognitive factors. Each of the 20 alternate forms consists of two pages of random numbers. Twenty-five rows of 30 digits on each page are presented/with the first number of each row circled. The task is to cross out each number in a row that is like the circled number at the beginning of the row. Two forms were divided in half and each half was used for each of four practice trials and ten of the alternate forms of two pages

were used  
each/for the ten experimental trials. The score was the number of rows marked entirely correctly.

#### Procedure

Subjects were told that the test measured perceptual speed. All subjects were given four 30 second practice trials separated by  $1\frac{1}{2}$  minute intervals. They were told to "do as many rows as they could" on each. After each practice trial, subjects were given the correct answers for each row.

The subjects were then told that there would be 10 experimental trials, each two minutes in length with a three minute rest interval between trials. At this point the experimental instructions were introduced.

High Standard (+8) Group (N=16). Subjects in this group were told that they should try to beat a score (in terms of number of rows entirely correct) on each experimental trial equal to the sum of the number of correct rows on the four practice periods (which they had just computed) plus 8. Subjects were told that this standard was "what we considered to be a successful performance on the task" and "somewhat above the average performance" for college students of their initial ability. They then wrote their standard at the bottom of the second page of each of the ten forms. At the end of each trial subjects were given the correct answers and counted the number of rows marked correctly. They wrote this score next to their standard.

Improvement (+1) Group (N=14). This group was given the same instructions as the High Standard Group except that each subject's standard of success was given as one more than his best previous score.

Do Best Group (N=16). Subjects in this group were told to "do their best" on each trial. They were given the correct answers for all trials but they did not actually add up the total number of rows done correctly on either the practice or the experimental trials.

Low Motivation Group (N=11). This group was told that we were predominately interested in attitudes about the task (as measured by rating scales given after each trial; these results are not reported here) rather than performance. These subjects were given the answers for all practice trials, but computed their "total" practice score by summing their scores on the last three practice trials only. They were told that the average college student was able to work 3 less correct rows on each trial than his total score on the last three practice trials and were asked to work "at about this speed." These subjects were given the correct answers for all experimental trials but did not add up the total number of rows correctly done each time (in order to de-emphasize the importance of their scores).

Relaxed Group (N=15). This group was told that we were pretesting a test we were developing to measure perceptual speed. They were asked to work at a "comfortable pace" and not to push or tire themselves too much as we wanted to develop norms for college students working at a "normal rate of speed." Correct answers were given for all trials, but the subjects did not add up the number of rows done correctly on any of the trials (again in order to de-emphasize their scores). At the end of the experiment, all subjects completed detailed goal description questionnaires asking about their performance goals during the experiment. These questionnaires are shown in Appendices A and B.

#### Results

Before analyzing the data, all subjects' work sheets were scored by the experimenter even for subjects who had previously scored their own (though few discrepancies in the latter cases were found).

#### Analysis by Experimental Condition

The subject's improvement score was the dependent variable. This score

was obtained by subtracting the subject's mean score on the 10 experimental trials from his total practice trial score (for the four practice trials combined). An F-test was performed on the mean improvement scores of the five experimental groups. To control for differences in initial (practice) performance, an analysis of covariance design was used. The differences between the five improvement means was not significant ( $F=1.75$ ), nor were there any significant differences between any pair of means. This suggested that the subjects were not pursuing their assigned goals, since the assigned goals differed considerably in difficulty (for groups where this could be computed). The High Standard (+8) subjects reached or exceeded their assigned goal on only 3% of the trials, whereas the Low Motivation subjects exceeded their "suggested" standard on 100% of the trials. The Improvement (+1) subjects reached or beat their assigned goal 28% of the time.

An examination of the goal descriptions filled out after the experiment indicated that, in fact, most subjects were not pursuing their assigned goals. Only 50% of the High Standard subjects stated that they were trying to reach or exceed their assigned goal; only 36% of the Improvement group and 25% of the Do Best subjects indicated that they were pursuing their assigned goals. Less than 7% of the Relaxed subjects and none of the Low Motivation subjects were trying for their suggested goals.

Due to this lack of within-group homogeneity in goals, it was decided to re-analyze the data according to the subjects' a posteriori goal descriptions.

#### Analysis by A Posteriori Goal Descriptions

The reclassification of subjects according to their a posteriori goal descriptions yielded approximately five major categories, as described below.



High Standard Group (N=9). This group was composed entirely of those subjects from the High Standard (+8) group who indicated that they were trying to reach or exceed the goal of eight more correct rows than they had completed on the four practice trials.

High Improvement Group (N=16). Subjects in this group stated that they were trying to beat their best previous score by one or more rows on each trial.

Low Improvement Group (N=12). This group consisted of those subjects who were either trying to equal their best previous score (N=4), trying to beat their immediately previous score (N=6), or those subjects trying to equal their immediately previous score (N=2). These groups could have been considered separately, but they were combined in order to yield a larger N. However, they all have in common the fact that their improvement goals were less difficult than those of the High Improvement subjects.

Do Best Group (N=22). The largest a posteriori group was composed of subjects who indicated they were trying to "do their best" without regard to their assigned standard or to their previous scores.

Miscellaneous (Low Motivation) Group (N=9). Subjects who stated they were working with "little effort," "just trying to complete the task," or "working mainly for accuracy," etc. were placed in this group. These subjects had in common the fact that they appeared not to be highly motivated and had no specific (quantitative) goal in mind.

Of the original 72 subjects, four were dropped from the analysis at this point. Three of these subjects indicated that they were competing with other subjects rather than trying for one of the above goals. No justification could be found for including them in any of the above a posteriori goal groups, particularly since the scores of the people they were competing

with were not known. A fourth subject stated that he was trying to beat his immediately previous score some of the time. There was no way to determine how frequently this subject was pursuing the goal of trying to beat his immediately previous score or how often he was just working along, so he too was dropped.

The mean improvement scores for each of the a posteriori goal groups broken down into subgroups according to original experimental condition are presented in Table 1. It is evident that the relative mean improvement scores of each a posteriori goal subgroup are reasonably consistent across experimental conditions, even without correcting the means for initial (practice) scores. The High Improve and Do Best subgroup means are consistently higher than those of the Low Improve and Miscellaneous (Low Motivation) subgroups. Thus, it seemed permissible to combine the goal subgroups across conditions. The overall means of each a posteriori goal group are shown in the last column of Table 1.

An F-test was performed on the mean improvement scores of these five groups. Again, because of differences between the groups in initial level of performance, an analysis of covariance design was used. This analysis yielded an F-ratio of 7.28 ( $p < .001$ ; d.f. 4;62). t-tests on the differences between pairs of means corrected for initial performance level were also performed. These t-values are shown in Table 2.<sup>2</sup> No significance labels have been placed on the t-values, as they would be meaningless in view of the large number of tests made (and the unequal *N*'s precluded a Tukey multiple comparisons test). These values are useful simply to show the relative size of the differences between the various means. In general, the High Standard, High Improve and Do Best groups showed the highest improvement scores and the Low Improve and Miscellaneous (Low Motivation)

groups showed the lowest improvement scores.

## Experiment II

### Method

#### Subjects

The subjects were the same as those used in the previous study. The two experiments were run consecutively in the order discussed.

#### Task

The task was simple addition, modeled on test N-1 from Factor N (Number Facility) in French, Ekstrom, and Price's (1963) reference kit for cognitive factors. The original test comes in two parts, each part consisting of sixty problems with two minutes allowed per part; each problem consists of three positive one or two digit numbers which must be summed. Eight more parts of sixty problems each were constructed using random number tables, thus making ten two-minute parts. Four additional parts of 30 problems each were constructed in the same manner for use as practice problems.

#### Procedure

All subjects were told that the purpose of the experiment was to study the development of adding skill, the relationship between early and late performance and attitudes toward the task. All subjects were given four 30-second practice trials on which they were told to "do as many problems as they could." After each practice trial (with one exception to be described below) they were given the correct answers to each problem. The interval between practice trials was 45 seconds.

All subjects were then told that there would be 10 two-minute experimental trials separated by a two-minute rest period, during which, among other things, they would be given the correct answers to each problem. At this point the experimental instructions were introduced.

High Standard (+8) Group (N=15). Each subject in this group was told that he should try to beat a score (in terms of total number of problems correct) on each experimental trial equal to the sum of the number correct on his four 30-second practice trials (which he had just computed) plus 8. He was told that this standard was "what we considered to be a successful performance on the task" and represented "somewhat above the average performance" for students of his initial ability. The subject then wrote the number representing his standard at the bottom of each experimental trial sheet. At the end of each trial, the correct answers were read, and the subject counted the number he got correct and wrote this score on the sheet next to his standard.

Low Standard (-6) Group (N=16). This group was given the same instructions as the High Standard group except that each subject's standard of success was given as 6 less than his total practice trial score.

Do Best Group (N=16). This group was told to "do their best" on each trial. In addition each subject set a level of aspiration ("the score you will try for") before each trial. It was suggested to the subjects that they set their levels of aspiration each time so that they were "neither too hard nor too easy." At the end of each trial, the obtained score was written next to the level of aspiration for that trial before the new aspiration level was set.

Incentive Group (N=14). This group was told they would be paid according to the number of problems they got correct on the ten experimental trials (the implicit goal being to "make money"). The incentive was set so that if they got the same number right as the Do Best subjects (who were paid for participation) they would get paid (on the average) the same amount (though the subjects did not know the incentive was set in this manner). These

subjects also set a level of aspiration before each trial and wrote their obtained score next to their aspired score after each trial.

Low Motivation Group (N=11). These subjects were not given the correct answers to the first practice trial, so when computing their "total" practice scores, they used their scores on the last three practice trials only. They were told that the average subject was able to get three less correct on each trial than the total score on the last three practice trials and to "go along at about that speed." These subjects were given the correct answers to the problems between experimental trials, as with the other groups, but they did not add up the number they got correct each time, nor did they set any explicit levels of aspiration. The reason for the latter omission was that it was felt that goal-setting itself might induce high motivation (e.g., see Fryer, 1964) whereas this group was supposed to be a low motivation group.

At the end of the experiment all subjects completed detailed goal description questionnaires asking them about their performance goals during the experiment. These questionnaires are shown in Appendices B and C.

#### Results

For the analysis, all subjects' work sheets were scored by the experimenter, even for subjects who were told to score their own (though few discrepancies in the latter cases were found).

#### Analysis by Experimental Condition

The dependent variable was the difference between the subject's score on all four practice trials combined and his mean score on the ten experimental trials. These scores will henceforth be referred to as improvement scores. The mean improvement score for each experimental group was calculated and an F-test was performed on the improvement means of the five groups.

An analysis of covariance design was used to control for initial level (which correlated  $-.61$ , within groups, with improvement). No significant differences were found between the five means ( $F=1.81$ ) nor between any pair of means. If all subjects had been pursuing their assigned goals, performance differences should have emerged, since the difficulty of reaching the assigned goals differed markedly for the five groups. The High Standard (+8) subjects were only able to beat their assigned standard on 1.3% of the trials; the Low Standard (-6) subjects, however, beat theirs 85% of the time, and the Low Motivation beat their "suggested" standard 99% of the time. The Incentive and Do Best groups beat their levels of aspiration 34% and 39% of the time, respectively, indicating that their self-set goals were of moderate difficulty.

However, an examination of the a posteriori goal descriptions indicated that most subjects were not pursuing their assigned goals nor were they necessarily pursuing their own levels of aspiration. Only 40% of the (+8) group was trying to "reach or just beat" their assigned goal; less than 19% of the (-6) group were trying to "reach" their goal. None of the Low Motivation subjects were trying for the "suggested" goal of their score on the last three practice trials (combined) minus 3. Fifty-seven percent and 62% of the Incentive and Do Best subjects, respectively, were trying to "reach" or "exceed" their levels of aspiration, and only 31% of the Do Best subjects were trying to "do their best" regardless of their goals. This suggests that subjects will be less prone to accept or try for goals that are too hard or too easy than goals which are more moderate in difficulty (such as ones they are likely to set themselves), although a surprising number of subjects who set their own goals denied trying for them.

Due to this lack of within-group homogeneity in goals, the data were

reanalyzed according to the a posteriori goal descriptions.

#### Analysis by A Posteriori Goal Descriptions

(+) and (-) Goals. The most immediately obvious way of reclassifying the subjects into a small number of homogeneous groups seemed to be according to whether the subject was trying for a score which was higher or lower than his score on the four practice trials (combined). Subjects who appeared to be trying for a score higher than their total practice score were classified as (+) subjects, and those who appeared to be trying for a score lower than their total practice score were classified as (-) subjects. The classification was done as follows: (a) in the +8 group, subjects who were trying to "reach", "exceed", or "get close" to the assigned standard were classified as (+); in addition, the one subject who was trying to "improve" each time was classified as (+), since the initial standard for improvement was presumably his total practice score. Two subjects who were trying for "accuracy" rather than speed were classified as (-) as this suggested less than maximal motivation. (b) In the -6 group, subjects who were trying to "exceed" the assigned standard by more than six points were classified as (+). Subjects who were trying to "reach", or "exceed the standard by less than six points" were classified as (-) along with one subject who was trying to "get through" (but not "get right") as many problems as possible. (c) In the Do Best group subjects whose mean level of aspiration on the ten experimental trials was greater than their total practice score and who said they were trying to "reach" or "exceed" these goals or to "improve" were classified as (+); subjects whose mean level of aspiration was less than their practice score and who said they were trying to "reach" their goals were classified as (-). (d) The same procedure was followed in classifying the Incentive subjects as for the Do Best subjects.

In the Low Motivation group, classification was more difficult since none of the subjects said they were trying for the suggested standard and no explicit levels of aspiration were set. One subject was trying to beat his practice score (for the last three practice trials) by 13, which would put him above his total practice score, and was classified as (+). One subject who had "no goal" was classified as (-). Seven subjects had a goal of "improve over best previous score." It might be argued that a goal of improvement should lead to a higher score for all subjects than their total practice score, so that all improvement subjects should be classified as (+). However, this ignores two things: (1) the original starting point (the original score taken as the standard for improvement), and (2) task difficulty. The Low Motivation subjects had knowledge of their score for only the last three practice trials; if they used this total as a starting point for improvement, this would have represented a standard anywhere from two to nine points (depending on the individual) below their total practice score, thus suggesting a very low overall standard. Since the expected mean improvement for all subjects was -0.88, a standard this low would be expected to lead a considerably lower mean score than their total practice score. However, there was no means of determining from the data available whether or not subjects took their practice scores (for last three trials) as their original standard or not. Thus it was decided to take the subject's score on the first experimental trial as the initial standard for improvement: if this score was two or less points under his total (four trial) practice score, he was classified as (+), since it implied that his initial standard for improvement was close to his total practice score; if his score on the first trial was three or more points under his total practice score, he was classified as (-). This put four subjects in the (+) group and three in the (-) group. If there was any bias



in these placements, it was probably in putting too many subjects in the (+) group. However, slight changes in the above criteria would not have yielded substantially different results.

All subjects in all conditions who indicated they were "doing their best" with no reference to their previous scores, assigned goals, or aspiration levels were temporarily removed from the analysis, though they will be dealt with subsequently. These subjects will henceforth be referred to as Do Best (ap), for "a posteriori", to distinguish them from subjects in the Do Best experimental condition.

The mean improvement scores of each sub-group described above are shown for each experimental condition in Table 3. In all cases the mean improvement scores of all (+) subjects combined are higher than those of all (-) subjects combined in the same experimental condition; in all but one case the means for all (+) subjects combined in each condition are positive, and in all cases the means for all (-) subjects in each condition are negative. The only discrepancy from the overall pattern of note for any individual sub-group is the high negative mean of the "get close" subjects in the (+8) condition. Since these subjects did not indicate just how close they were trying to get to the standard, putting them in the (+) group was giving them the benefit of the doubt.

In view of the consistent difference in the means of the (+) and (-) groups across conditions, it seemed permissible to combine all groups for purposes of analysis. The mean improvement scores of all (+) subjects combined and all (-) subjects combined are shown at the bottom of Table 3. These means are significantly different at the .001 level after correcting for initial (practice) scores ( $t=4.19$ ; 49 d.f.).<sup>3</sup> Fifty-three percent of the (+) subjects changed in a positive direction, while one hundred percent of the

(-) subjects changed in a negative direction. The chi-square value for relative frequency of direction of change is 11.69 which is significant at the .001 level (1 d.f.).

Further Breakdown of (+) and (-) Subjects. The (+) and (-) goals described above could be further broken down as follows: a (++) group was constructed by taking those subjects in the High Standard (+8) group who indicated they were trying to "reach" the assigned standard and any subjects in the Do Best and Incentive experimental conditions whose mean level of aspiration was 8 or more points above their total practice score. Similarly a (--) group was constructed by taking those subjects in the Low Standard (-6) condition who said they were trying to "reach" the Low Standard and any Do Best or Incentive subjects whose mean level of aspiration was 6 or more points below their total practice score. The means of these new groups in addition to the means of the original (+) and (-) groups with the (++) and (--) subjects removed are shown in Table 4. The means as shown are not corrected for initial practice scores which favor the differences, but the mean difference between the (++) and (--) groups is significant at the .01 level after correction ( $t=3.31$ ; 10 d.f.), as is the mean difference between the remaining (+) and (-) subjects ( $t=3.18$ ; 37 d.f.).

Do Best (ap) Subjects. The mean improvement score of all Do Best (ap) subjects is shown at the bottom of Table 3. The mean change for this group is very similar to that for the (+) group but significantly greater than that of the (-) group.

The breakdown of the Do Best (ap) subjects by experimental condition is shown in Table 5. It is evident that the five Do Best (ap) Incentive condition subject showed a higher mean improvement than the other groups and the five Do Best (ap) subjects in the Do Best condition had a lower mean.

It will be recalled that all Do Best and Incentive condition subjects were required to set specific levels of aspiration for each trial. It is possible that these subjects were actually trying for their goals even though they claimed that they were just trying to "do their best". To test this possibility these ten subjects were broken down into a (+) group and a (-) group according to whether or not their mean level of aspiration was above or below their total practice score. The results are shown at the bottom of Table 5. The mean improvement scores of these Do Best (ap) (+) and (-) sub-groups are significantly different at the .05 level, after correcting for initial practice scores which favored the difference. This suggests that subjects who are required to set explicit levels of aspiration do pursue these goals even though they claim they are "doing their best" with no reference to their goals.

Parallel inference<sup>s</sup> could not be made, of course, for the Do Best (ap) subjects in the remaining three experimental conditions since these subjects did not set explicit performance goals. Thus they may be considered to be "genuine Do Best (ap)" subjects. It can be seen that the mean improvement score of these remaining Do Best (ap) subjects (bottom of Table 5) is directly between those of the (+) and (-) groups shown in Table 4.

Finally Table 6 summarizes all the findings of this study by showing how each original a posteriori goal group was broken down into goal sub-groups and the N's and mean improvement scores of each of these groups, as well as the results of relevant t-tests.<sup>4</sup> It is clear that each further breakdown of the goal groups resulted in the formation of more homogeneous (with respect to performance) sub-groups.

#### The Variability of Do Best Goal Subjects

Of all the goals delineated and discussed in the present investigations, the Do Best goal is perhaps the most vague. It is distinguished from goals

such as "improve," "reach the high standard," "reach the low standard," etc. by the fact that it does not involve an explicit performance standard or goal. Thus one might expect that performance under this goal would be determined by numerous extraneous factors such as implicit goals, the mood of the moment etc. which might be different for each person. If this were true, we might expect that the variability in performance<sup>6</sup> of Do Best subjects to be greater than that of subjects with more determinate goals. The above reasoning was not entirely deductive but was developed in part an examination of the Do Best as compared to the other subjects in the second study. We then went back and made the same comparisons for the first study, and then back to two previous studies we had done involving "Do Best" goals. The findings were highly consistent across studies, and are shown in Table 7. Each F-ratio represents the ratio of the  $s^2$  for a Do Best group in comparison to the  $s^2$  of each other goal group in the same experiment. The striking aspect of these findings is that the Do Best group variance is greater than that of the other goal groups (i.e. greater than 1) without a single exception in 14 different comparisons. Although the majority of these F-ratios are not significant, the consistency of the direction of difference is far beyond the chance level. On the average the Do Best group variance is nearly double that of other goal groups.

#### Discussion

Before discussing the specific findings, it is relevant to ask to what degree the goal descriptions of our subjects could be considered valid. We have shown already (by the use of statistical controls) that they were not an artifact of the subjects' initial scores, but one might ask whether these descriptions could simply be post hoc rationalizations of their performance. In other words, couldn't the subjects have performed in a certain manner and

then invented goals that were congruent with their performance? A number of considerations argue against this. First, this would not explain why people would invent a goal of "Do Best," since such a goal would not "explain" anything. Second, there were people who did not reach their stated goal at all (or even get close to it), and there were others who exceeded it a great deal (i.e. there were numerous individual discrepancies between stated goals and performance). Third, among people who set levels of aspiration, there were individuals who set very high or very low goals (in comparison to their practice scores) on the very first trial before they could have known how well they would do. Finally, previous studies in which similar goals were successfully manipulated in advance by instructions (e.g. Mace, 1935; Locke, 1965) yielded comparable results. The evidence suggests that at least in the short term, most subjects will not inevitably try to rationalize failure to reach a goal by inventing a lower one after the fact. Though there may have been individual instances of rationalization, it is argued that this cannot explain the results as a whole. In addition, it should be pointed out that there are cases where an individual genuinely changes his goal during an experiment; for instance, an individual who cannot reach a hard goal might change to a lower goal, but this goal will in turn actually influence his subsequent performance. To determine these kinds of effects, of course, goal questionnaires would have to be given before (and after) every trial.

At the most general level the results of these experiments suggest that performance goals that differ both quantitatively and qualitatively can be systematically related to level of performance on performance tasks even under maximally disadvantageous conditions (i.e. where the time periods are extremely short and the changes in performance over time extremely small).

Two factorially independent tasks were used in the present study indicating that the results are not restricted to a single task. The finding that qualitative as well as quantitative goals could be related to performance level complements previous work using learning tasks, (Mace, 1935; Locke and Bryan, 1965b) and supplements previous work using performance tasks where only quantitatively different goals were used (Locke, 1965).

At the more specific level, previous findings by Locke (1965) and Locke and Bryan (1965b) that high or hard performance goals resulted in a higher level of performance than low or easy performance goals were replicated. In the first experiment reported here, the High Standard group improved more than the two low motivation groups (Low Improve and Misc. Low Motivation). In the second experiment the (++) and (+) groups improved more than the (--) and (-) groups, respectively. The same was true of the (+) Do Best (ap) subjects as compared with the (-) Do Best (ap) subjects. The difference in the difficulty of the high and low goals was attested to by the differences in empirical probability of beating the goals (where this could be calculated). In the first study the High Standard and High Improve subjects were able to reach or beat their stated goals only 5% and 21% of the time respectively, while the Low Improve group was able to beat theirs 39% of the time. In the second study the (++) group beat their standard only 3.8% of the time while the (--) group beat theirs 62% of the time.

In relation to the hard and easy goal subjects, the Do Best (ap) subjects in the present studies performed significantly better than the various low motivation groups (Low Improve group in study I; (-) group in study II) but not significantly worse than the high motivation groups (High Standard and High Improve groups in study I; (+) group in study II). This points out a possible difference in the effects of hard and "Do Best" goals in learning

and performance tasks. Previous studies by Mace (1935), Locke and Bryan (1965a,b) with learning tasks found a Hard Standard group to be consistently superior to a Do Best group. All subjects in these latter experiments improved (and undoubtedly had at least an implicit goal to improve) but the High Standard subjects improved at a faster rate. In the performance tasks used here where improvement was almost non-existent (and with short trial periods) extra hard goals may not have been more effective than Do Best goals because extra effort of the type presumably induced by hard goals may not "pay off" on such tasks. In other words, given a reasonably high degree of motivation and a task where improvement is impossible, the only way to go is "down," thus the difference between both the High Standard and Do Best goals and the various low motivation goals. High goals on performance tasks such as those used here apparently keep performance rate from falling, but do not increase it appreciably.

However it should be made clear that the effects of high performance goals might be considerably greater in comparison to Do Best goals where the trial periods are longer, i.e. where persistence over long periods can influence performance. There was little or no opportunity for persistence to be a factor in the present studies, consisting as they did of only 10 2-minute trials each.

It was stated earlier that less difference between Do Best and Improve goals was to be expected on performance tasks like those used here than on learning tasks. In the second study it was noted that the meaning of the "Improve" goal was somewhat ambiguous for the subjects who stated it; but computing the mean improvement score of all subjects who stated they were trying to improve yields a mean improvement score very close to that of the Do Best (ap) group, thus supporting the hypothesis. However, the results

of our first study showed that the difference between the Improve and Do Best goals depends upon the rate of improvement tried for. There was virtually no difference between the High Improve group and the Do Best group, but there was a large difference between the Do Best group and the Low Improve group in favor of the former.

Again with respect to the Do Best goal, the results of the second study suggested that under conditions where specific levels of aspiration are set, subjects who claim they are trying to "do their best" with no reference to goals or previous scores actually do try for their goals. Perhaps if all Do Best subjects could be questioned more closely it would be found that many of them do have specific goals, at least at the implicit level.

Finally, there are practical implications of the finding that the Do Best goal results in greater variation in performance than is the case with other goals. It suggests that in experiments where maximally homogeneous groups are wanted, the assignment of a Do Best goal may not only lead to ignoring the goal completely and the setting of new goals in many cases, (though this is also true of other kinds of goals) but for those who do accept the goal, it will lead to relatively wide variations in performance and thus attenuate the significance of differences between this group and other groups. Since minimum within-group variation is ordinarily desirable in psychological experiments, the use of Do Best instructions to produce Do Best goals may not always be a wise procedure.

This brings us to the problem of goal acceptance in general. A remarkably small number of subjects in these studies accepted the goals assigned to them. Although such widespread individual goal rejection and substitution may not be typical in other types of experiments, the phenomenon is probably not uncommon in psychological research. Intensive studies in the area of verbal learning and awareness have found wide variations in subjects' hy-



potheses and intentions to be the rule rather than the exception (e.g. see Spielberger, 1965, for a review of this work). As Spielberger points out, the epistemological biases of behaviorists which have permeated the field have led to an implicit (and sometimes explicit) "robot-theory" of experimental manipulation. It is usually assumed that the subject will do exactly what he is told and that what he thinks, the conclusions he comes to and the judgments he makes during the experiment will have no important bearing on his experimental behavior. This has led to a relative neglect of cognitive variables in motivation research and to an nearly exclusive focus on "instructions" and "stimuli" as independent variables. Suffice it to say that not a single significant difference would have been found between any pair of groups in either of the two studies reported here, if "instructions" had been the only independent variable used. And even more important for the purposes of science, without questioning the subjects, we would not have known why the instructions did not work. Of all the criticisms that could be made of not trying to find out what subjects are thinking and trying to do, this is perhaps the most important, because failure to question subjects (or failure to question them properly) can lead to false conclusions (as it did in the verbal learning and awareness area; see Spielberger, 1965) or to type II errors, as would have happened in the present case. As an example, we can take the finding that the group in the second study under the Incentive plan did not show greater improvement than the other groups. Without questioning the subjects, we would have had to conclude that "incentives don't work under these conditions," but we would not have known why they didn't work. On the basis of the goal descriptions, we can suggest that the reason they did not work was that the Incentive subjects did not set high enough goals in relation to the other groups. Supporting this interpretation is a previous study of choice

behavior by Locke and Kendall (1965) which found that the influence of monetary incentives on choice behavior was a direct function of the goals the subjects developed in response to such incentives.

A similar explanation was offered by Locke and Bryan (1965b) to explain the fact that explicit knowledge of total score did not automatically result in higher performance than was the case without such knowledge. Again performance was found to depend upon what goals the subjects set in response to such information. The same was true in the present studies; it will be recalled that some experimental groups in each study had specific knowledge of their total scores (though they all had general knowledge of how well they were doing) and others did not, yet no differences between experimental groups were found. The important thing was what goals the subjects set on the basis of the knowledge they did have. It is likely that similar findings could be obtained to explain the effects of other types of "instructions," "extrinsic motivators," and "stimuli."

#### Footnotes

- 1 This research was supported by Contract Nonr 4792(00) between the American Institutes for Research and the Office of Naval Research. The opinions expressed do not necessarily represent those of the Department of the Navy.
- 2 The correction factor for initial (practice) score for the means for all the t-tests was the slope ( $b_{xy}$ ) as obtained in the covariance analysis by a posteriori goals using 68 subjects. No corrections were made of the within group sums of squares in computing the t-values, so the values shown may under-estimates of the true t-values.
- 3 All corrections of means (for initial practice scores) in this study were made using the slope ( $b_{xy}$ ) as computed in the preliminary covariance analysis by experimental conditions (and including all 72 subjects).
- 4 All possible t-tests were not performed on these data due to the problem of multiple comparisons.

#### References

- Festinger, L. A theoretical interpretation of shifts of level of aspiration. Psychological Review, 1942, 49, 235-250.
- Frank, J. D. Recent studies of level of aspiration. Psychological Bulletin, 1941, 38, 218-226.
- French, J. W. Ekstrom, R. B., and Price, L. A. (Eds.) Kit of reference tests for cognitive factors, Princeton, N. J.: Educational Testing Service, 1963.
- Fryer, F. W. An evaluation of level of aspiration as a training procedure, Englewood Cliffs, N. J.: Prentice Hall, 1964.
- Lewin, K. Psychology of success and failure, in C. L. Stacey and M. F. DeMartino (eds.) Understanding human motivation, Cleveland: Allen Inc., 1958, 223-228.
- Locke, E. A. The relationship of intentions to level of performance. Journal of Applied Psychology, 1965, (in press).
- Locke, E. A. and Bryan, J. F. Cognitive aspects of psychomotor performance: the effects of performance goals on level of performance. American Institutes for Research (unpublished), Silver Spring, Md., 1965. (a)
- Locke, E. A. and Bryan, J. F. The effects of goal setting, rule learning and knowledge of results on performance. American Institutes for Research, Silver Spring, Md., 1965, (unpublished). (b)
- Locke, E. A. and Kendall, L. M. The relationship of goals, plans, and intentions to choice behavior. American Institutes for Research, 1965 (unpublished).
- Mace, C. A. Incentives: some experimental studies, Industrial Health Research Board (Great Britain), 1935, Report No. 72.
- Moran, L. J. and Mefferd, R. B. Repetitive psychometric measures. Psychological Reports, 1959, 5, 269-275.

- Rotter, J. B. Level of aspiration as a method of studying personality: I, critical review of methodology. Psychological Review, 1942, 49, 463-474.
- Siegal, S. & Fouraker, L. E. Bargaining and group decision making, New York: McGraw-Hill, 1960, Pp 61-69.
- Spielberger, C. D. Theoretical and epistemological issues in verbal conditioning. in S. Rosenberg (Ed.) Directions in psycholinguistics, 1965 (mimeo draft).

Table 1

N's and Mean Total Improvement Scores for Goal Sub-Groups by Experimental Condition  
(Experiment I)

A Posteriori Goal Group	Measure	Experimental Condition					All Subjects	
		High Std. (+3)	Improve (+1)	Do Best	Low Motivation	Relaxed	Mean	N
High Standard	Mean <sup>a</sup> n	9 +1.88	-- --	-- --	-- --	-- --	+1.88	9
High Improvement	Mean n	1 +0.50	5 +0.90	2 +2.45	6 +2.18	2 +1.10	+1.58	16
Low Improvement	Mean n	-- --	3 -1.20	3 -0.70	-- --	6 +0.78	-0.08	12
Do Best	Mean n	5 +1.40	5 +2.96	4 +0.50	5 +0.20	3 +2.60	+1.48	22
Misc. (Low Mot.)	Mean n	1 +0.10	1 +0.20	5 +0.48	-- --	2 -0.85	+0.11	9
All Subjects	Mean N	16 +1.59	14 +1.14	14 +0.51	11 +1.28	13 +0.99	+1.10	68

<sup>a</sup> These means are not corrected for initial (practice) scores.

Table 2

t Value for Mean Differences in Improvement Scores of  
A Posteriori Goal Groups  
(Experiment I)

	High Standard	High Improve	Low Improve	Do Best	Misc.(Low Mot.)
High Standard	----	1.26 <sup>a</sup>	2.84	<1	2.40
High Improve		----	2.56	<1	1.79
Low Improve			----	2.44	<1
Do Best				----	<1
Misc.(Low Mot.)					----

<sup>a</sup> All t values are based on means corrected for initial (practice) scores.

Table 3

N's and Mean Improvement Scores For Goal Sub-groups by Experimental Condition  
(Experiment II)

A Posteriori Goal Group	Experimental Condition				
	(+8) High Std.	(-6) Low Std.	Do Best	Incentive	Low Motivation
Subjects Classified as (+)	(6) Reach Std. (+0.90) <sup>a</sup> (4) Get Close (-3.78) (1) Improve (+3.00)	(1) Exceed Std. by > 6 (+2.70)	(9) Reach or exceed $\overline{IA}$ which is > P <sup>c</sup> (-0.05) (1) Improve with $\overline{IA}$ > P (+1.80)	(7) Reach or exceed $\overline{IA}$ which is > P (+0.13) (1) Improve with $\overline{IA}$ > P (+0.50)	(4) Improve, if trial 1 score within 2 of P (+1.58)
Total N: Overall Mean	(11)  (-0.61)	(1)  (+2.70)	(10)  (+0.13)	(8)  (+0.18)	(4)  (+1.58)
Subjects Classified as (-)	(2) Misc. (-3.15)	(3) Reach Std. (-5.17) (5) Exceed Std. by < 6 (-2.86) (1) Misc. (-4.60)	(1) Reach or exceed $\overline{IA}$ which is < P (-5.90)	(1) Reach or exceed $\overline{IA}$ which is < P (-10.80)	(3) Improve, if trial 1 score more than 2 below P (-2.77) (1) Misc. (-3.90)
Total N: Overall Mean	(2)  (-3.15)	(9)  (-3.82)	(1)  (-5.90)	(1)  (-10.80)	(4)  (-3.05)
Total N (+) : 34 Total N (-) : 17 Total N (A Posteriori Do Best) : 21					Mean (+) : +0.15 Mean (-) : -4.09 Mean (A Posteriori Do Best) : +0.00
					$t = 4.19^d$ p < .001 $t = 2.73^d$ p < .01

<sup>a</sup> numbers in parentheses to left of each goal indicate number of subjects pursuing that goal;  
numbers in parentheses to right indicate mean improvement scores of subjects pursuing that goal.  
These means are not corrected for differences in initial (practice) scores.

<sup>b</sup>  $\overline{IA}$  refers to the mean level of aspiration on the ten experimental trials.

<sup>c</sup> P refers to the total score on the four practice trials.

<sup>d</sup> After correction for initial (practice) scores.



Table 4

N's and Mean Improvement Scores of (++) , (+) , (-) and (--) Subjects  
(Experiment II)

Measure	A Posteriori Goal Group			
	(++) <sup>a</sup>	Remaining (+)	Remaining (-)	(--) <sup>b</sup>
N	8	26	13	4
Mean <sup>c</sup>	+1.64	-0.31	-3.33	-6.58

<sup>a</sup> Subjects who were trying to beat their practice scores by 8 or more points were classified as (++) .

<sup>b</sup> Subjects who were trying to beat their practice scores by -6 points or worse were classified as (--).

<sup>c</sup> These means are not corrected for initial (practice) scores.

Table 5  
Mean Improvement Scores of Do Best (ap) Subjects  
(Experiment II)

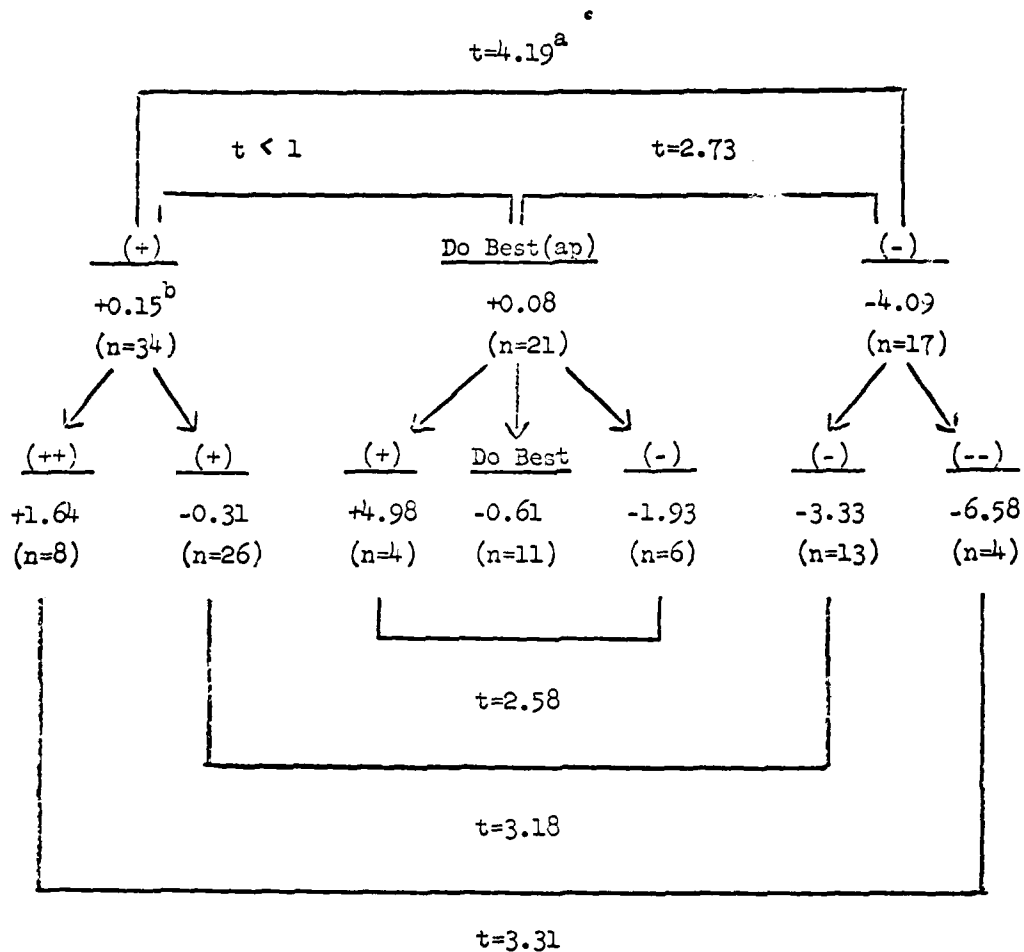
Measure	Experimental Condition					All Do Best ( <u>ap</u> ) Subjects
	High Standard (+8)	Low Standard (-6)	Low Motivation	Incentive	Do Best	
N	2	6	3	5	5	21
Mean	-1.10	-1.10	+0.70	+3.50	-1.84	+0.08
	Mean: -0.61 (N=11)			Mean for (+) <sup>a</sup> : +4.98 (N=4) Mean for (-) <sup>a</sup> : -1.93 (N=6)] $t=2.58^b$ $p<.05$		

<sup>a</sup> Subjects whose mean LA exceeded their total practice score were classified as (+); those whose mean LA was less than their total practice score were classified as (-). See text for further explanation.

<sup>b</sup> Adjusted for initial (practice) score.

Table 6

Breakdown of A Posteriori Goal Groups: Mean Improvement Scores and N's  
(Experiment II)



<sup>a</sup> All t's are based on means corrected for initial (practice) scores.

<sup>b</sup> Means as shown are not corrected for practice score.

Table 7  
Comparison of Variance of Subjects with Do Best Goal  
with Variance of Subjects with Other Goals  
(Four Experiments)

Study	$s^2$ of Do Best Subjects/ $s^2$ of:	F	d.f.	p
Locke & Bryan, 1965a:	High Standard S's	2.64	13;13	.05
Locke & Bryan, 1965b:	High Standard S's	1.80	16;4	ns
	Improvement S's	1.15	16;23	ns
	Other (Low Mot.) S's	1.55	16;23	ns
Present Study Experiment I:	High Standard S's	1.01	21;8	ns
	High Improve S's	2.05	21;15	ns
	Low Improve S's	1.002	21;11	ns
	Other (Low Mot.) S's	2.85	21;8	ns
Experiment II:	(++) S's	2.09	10;7	ns
	(+) S's <sup>a</sup>	1.93	10;25	ns
	(-) S's <sup>b</sup>	5.75	10;12	.01
	(--) S's	1.22	10;3	ns
	(+) Do Best ( <u>ap</u> )	1.26	10;3	ns
	(-) Do Best ( <u>ap</u> )	1.43	10;5	ns
Mean F = 1.98				

<sup>a</sup> not including (++) S's

<sup>b</sup> not including (--) S's

## Appendix A

### Goal Questionnaire for High Standard, Improve, Do Best and Relaxed Experimental Groups in Experiment I

We would like to know something about the goals you were pursuing during the preceding experiment. In section A, check the one statement which describes your predominant goal. Please indicate what you were actually trying to do, regardless of the instructions you were given. Use section B to describe any goal you may have had which is not listed, or if at some time during the experiment you were not following the goal you checked in section A. Be sure to explain when during the experiment your goals changed, if at all.

#### A.

- ☐ a I was trying to reach or exceed the standard on every trial.
- ☐ a I was trying to reach or exceed the standard on \_\_\_\_\_ of the trials.
- ☐ a I was trying to get as close to the standard as possible, though I was not trying to beat it.
- ☐ I was trying to beat my best previous score by \_\_\_\_\_ row(s) each time.
- ☐ I was trying to at least equal my best previous score each time.
- ☐ I was trying to beat my immediately previous score each time.
- ☐ I was trying to equal my immediately previous score each time.
- ☐ I was trying to do my best each time; I paid no attention to the standard or my previous scores.
- ☐ I was not trying to improve at all; just tried to keep going at a reasonable pace.
- ☐ I was trying to go along with as little effort as possible.
- ☐ I was trying only to keep going at a fairly slow pace.
- ☐ I was trying to improve over the score I got on the first trial, but not after that.
- ☐ I was trying to give the illusion of working but actually I was hardly doing anything.

#### B. OTHER (Explain fully) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

<sup>a</sup>Note: only the High Standard group were given these items (in addition to the others).

Appendix B

Goal Questionnaire for Low Motivation Experimental Groups  
in Experiments I and II

I. Check the statement below which best represents the goal you had during the 10 experimental trials.

- ☐ I was just trying to go along with no particular goal in mind.
- ☐ I was trying to reach or just beat a score equal to my Practice<sup>a</sup> score minus 3 points each time.
- ☐ I was trying to reach or just beat my Practice<sup>a</sup> score each time.
- ☐ I was trying to beat my Practice<sup>a</sup> score by  points each time.
- ☐ I was trying to do the best I could with no reference to my previous scores at all.
- ☐ I was trying to improve my best previous score each time.
- ☐ Other: (explain)
- 
- 

<sup>a</sup>Note: the Practice scores for these subjects referred to their total score on the last three practice trials only.

Appendix C

Goal Questionnaire for High Standard, Low Standard, Do Best

and Incentive Experimental Groups in Experiment II

- I. Check the statement below which best represents the goal you had during the 10 experimental trials, with respect to the standards you were given or which you set for yourself. Fill in where necessary.

\_\_\_\_\_ I was trying to exceed the standard<sup>a</sup> each time by \_\_\_\_\_ points.

\_\_\_\_\_ I was trying to reach or just beat the standard<sup>a</sup> each time.

\_\_\_\_\_ I was trying to get as close to the standard<sup>a</sup> as I could each time, but was not trying to beat it.

\_\_\_\_\_ I was not trying for the standard<sup>a</sup> at all; I was just trying to do the best I could with no reference to my previous scores at all.

\_\_\_\_\_ I was not trying for any of the above goals; instead I was:  
(explain) \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

<sup>a</sup>Note: for the Do Best and Incentive Groups, the "Standard" was defined as their level of aspiration (the score they were trying for).